

# Theories of Evolution and Cultural Diffusion: The Dryad Repository Case Study for Understanding Changes in Organizing Information Practices

XXX

1st author's email address

1st author's affiliation

1st line of address

2nd line of address

Telephone number, incl. country code

## ABSTRACT

Digital networked and graphical technology is having a significant impact not only on the processes for creating, disseminating, and sharing information, but the practices for *organizing information*. These changes need to be studied and understood in order to keep pace of new developments and to consider how they may improve information organization. This paper considers theories of evolution and cultural diffusion for explaining and contextualizing the change in organizing information practices. The paper discusses these theories and their value for studying change. A case study focuses on the Dryad repository for data underlying published research in the field of evolutionary biology. This inquiry indicates that the theories of evolution and cultural diffusion are applicable to Dryad and reflective of change.

## Categories and Subject Descriptors

H.3 INFORMATION STORAGE AND RETRIEVAL

## General Terms

Design, Standardization, Theory.

## Keywords

Organizing information, Evolution, Cultural diffusion, Digital repository.

## 1. INTRODUCTION

Traditional physical barriers for creating, storing, and disseminating intellectual content are being eliminated by pervasive networked information technologies supporting instantaneous production and sharing of digital output in nearly any format. These developments have motivated the following seemingly significant changes in organizing information practices: 1. *an increase in access to a wide-variety of systems for organizing information*; 2. *ongoing development of new systems for organizing information—ranging from folksonomies to formalized ontologies*; and 3. *a remarkable growth in the diversity of individuals, well beyond traditional information professionals, using these systems for organizing information*.

One of the most visible places to observe organizing information change is in the social computing arena (e.g., Flickr, MySpace, and Connettea), where individuals tag for personal use, but also to contribute to a community's collective knowledge structure. Social computing spans non-academic and academic communities, and enables content creators and viewers to

represent and organize content via folksonomic tagging and through indexing by common attributes, such as *date*, *title*, and *creator (author)*. These approaches together can be grouped under the concept of metadata—structured data about data. Social computing and personal digital information applications are also supporting more formalize metadata schemes. For example many digital cameras and photo-posting sites like Flickr support the Exchangeable Image File Format (EXIF), a specification for the image files, and iTunes and Windows Media Player, among other music software applications, support the ID3 standard for MP3 audio files. (The ID3 acronym that was inspired by the decision tree algorithm developed by Ross Quinlan, although the acronym does not have an official translation in the context of MP3 metadata.) The burgeoning digital repository environment for scholarly output in multiple forms appear to have also cultivated significant change in organizing information practices, as many such repositories, Information request or even require that contributors (generally content creators) depositing their work, to help curate and create metadata describing their contributions.

The developments noted here feed O'Reilly's notion of Web 2.0 [1], and exhibit aspects of Library 2.0 [2], such as innovation and a participatory community. These developments provoke important questions about the changes in organizing information practices and the underlying systems, including the creation and use of knowledge organization systems (KOS). How might we explain or make sense of this change? What theoretical constructs may help us to understand the features and functionalities being presented in evolving information systems? Exploring the transformation in organizing information systems can help information science professionals understand change, and develop approaches to facilitating effective changes—leading to greater productivity. Moreover, specific to research, theory is crucial for identifying significant research questions for advancing our field.

This paper considers these questions via a case study approach and explores changes in organizing information systems in the context of *evolution* and *cultural diffusion*. The case study is presented with Dryad (<http://www.datadryad.org/>)—a digital repository for data underlying research publications in evolutionary biology and related fields. The case study interprets organizing information systems as schemes ranging from somewhat archaistic Folksonomies to formalized ontological structures. Information system refers to a framework supporting the acquisition/ingestion, representation, and search/retrieval of information objects. Transformation factors include 1. technological innovation, 2. the “networked” information infrastructure, 3. the enormous amount of information people

create and obtain daily, 4. the natural inclination to manage information so it can be found and effectively used, and 5. a growing and diverse population of people involved in organizing information.

The rest of the paper is organized as follows. The next section of this paper gives an overview on the theories of *evolution* and *cultural diffusion* as construct for studying change. The paper then introduces the Dryad project, reporting on the project's status and goals. Questions guiding an initial inquiry on the applicability theories of evolution and cultural diffusion are presented, and they are explored as a way to understand organizing information practices within Dryad. The paper concludes with brief remarks about next steps and the importance of exploring theory to understand change in organizing information.

## 2. Theories of Evolution and Cultural Diffusion

Proposing that the theories of *evolution* and *cultural diffusion* may help us understand changes in systems for organizing information seems a bit lofty, particularly without the support of empirical data. Despite this observation, there is value in first articulating that existing theories of "change" may in fact help us understand the transformation occurring in organizing information systems, and that observable evidence via a case study can provide grounding for further research. In other words, if this case study demonstrates the applicability of these theories, information science researchers may be encouraged to further explore these theories and be inspired to identify new and important research questions germane to our evolving field. This paper presents a logical first step by defining the theories of evolution and cultural and demonstrating changes supporting these theories observed while developing the Dryad repository

### 2.1 Evolutionary theory

The study of evolution is grounded in biological science and the idea that all living species have common ancestors. A simplistic example is family relationships, in which cousins have a common grandmother and inherit certain genes (e.g., the ability to have blue eyes). The evolutionary mechanism most familiar to the average citizen is *natural selection*—a process whereby a species inherits favorable characteristics over time for survival, popularized in Darwin's *Origin of the Species* [3]. *Adaptation* is another fairly familiar evolutionary concept. People living in extremely cold climatic regions (Alaska and Siberia) for thousands of years demonstrate adaptation via developing short round bodies and short arms and legs, and other characteristics, to minimize heat loss and endure extreme cold and wind. The physique of people living in such regions differs from their ancient ancestors who migrated to more arid and warmer climates.

The study of evolution is grounding in biology; it is an interdisciplinary field drawing from a wide range of scientific disciplines (ecology, genomics, paleontology, population genetics, physiology, systematics, and new biological subdisciplines such as genomics). Scientific study focuses on *phenotype*, the observable characteristics, traits, behaviors, and mapping a species morphological development; and *genotype*, the

genetic composition of an organism or cell. Our popular understanding of evolution, informed by Darwin and noted above, and is grounded more in phenotype, although not exclusively. Many disciplines have looked to evolution to develop their work, such as mathematics and computer science, whereby studying phenomena and developing genetic algorithms, allows for ruling out properties and deriving a single optimal solution [4].

A question to consider is can aspects of evolutionary theory explain changes in organizing information systems? Information systems are not living entities depending on food for sustainability, although they are operational systems with a lifespan. We study and develop information systems to support the life-cycle of digital information, and researchers have even defined a metadata life cycle [5]. Information systems connect to and are often related to other information systems and structures in the larger information infrastructure. For example, the World Health WHOSIS (WHO Statistical Information System) (<http://www.who.int/whosis/en/>) links to health-related statistical information found across the globe, and one can query the system to construct tables for any combination of countries. The linking and ability to integrate and develop new resources may represent a form of ancestry. These initial observations, it seems, warrant the exploration of evolutionary theory in relation to changes in organizing information systems.

### 2.2 Cultural Diffusion

Cultural diffusion can be defined as the spread of ideas, material objects, behaviors (eating, talking, dressing) from one society another [6]. Cultural diffusion is spurred by cultural domination or force, direct communication, and indirect means [7]. Cultural imperialism—forcing behavior, customs, or language of one culture on another is an example of cultural domination. It's difficult to see the applicability of domination for organizing information systems, although it may be possible to argue that domination exists if a vendor corners the market. What seems more probable are *direct* and *indirect* paths of cultural diffusion (among subcultures or communities) fostering change in organizing information systems.

The phenomenal growth in information technology over the last 30 years has allowed many *subcultures* to thrive and new communities to develop. Facebook—a social network web site developed initially for college students is one example of such a new community. There is clear evidence of *direct diffusion* (communication between information professionals and educators) as part of the development of many educational software packages, which, as a result, often include educational metadata standards and vocabularies for representing educational resources. Perhaps more a result of *indirect diffusion*, scientific communities are developing knowledge representation systems in the form of sophisticated ontologies for topics such as genetic sequencing and proteomics. Indirect diffusion is proposed here because the development of ontologies mimics, to some degree, the development of thesauri, although often scientific communities are not connected with the information and library science community.

Information/library scientists demonstrate a predilection for studying the spread of a topic and theories of diffusion. For example Bradford's law [8] provides a measure for examining how a "subject" is distributed (scattered) in the literature

according to a certain mathematical equation. This law can be used to study journals covering a discipline over a selected time period in order to understand the growth of a discipline. This law, and other bibliometric and informetric measures offer a view of diffusion. More precisely addressing diffusion, information and library science researchers have studied Rogers diffusion of innovation theory [9] in the context of topics such as collaboration with information systems [10], and the implementation and adoption of online searching [11]. These examples support an initial inquiry about cultural diffusion as a construct for explaining information organization changes, and motivate us to inquire and more formerly investigate if *ideas*, *objects*, and *behaviors* representing change in organizing information, in the context of information/library science practice can be explained, to any degree, by cultural diffusion.

### 3. The Dryad Repository

Dryad is a repository for data objects supporting published research in the field of evolutionary biology and related disciplines. Formerly known as DRIADE, the Dryad repository manages the life-cycle of heterogeneous digital data connected with the publication by facilitating data deposition (acquisition), preservation, discovery, sharing, use and reuse. The repository was launched via a collaboration involving the National Evolutionary Synthesis Center (NESCent) and UNC's School of Information and Library Science, Metadata Research Center <MRC>. Dryad seeks to balance a need for low barriers, inviting contributions from the wide range of scientists in evolutionary biology, with higher-level goals supporting computational data analysis for advancing evolutionary biology.

Organizing information is critical to the success of Dryad. A chief goal is to organize Dryad's objects, although approaches implemented in traditional organizing information systems (e.g., library catalog and indexes) are not fully suitable. Repository stakeholders have articulated a need and desire for search/retrieval capabilities and recognize that the use of some form of metadata standard and, more significantly, a vocabulary standard (e.g., a subject thesaurus) to help achieve these desired functions. Dryad differs from more traditional systems in that scientists are being required to contribute data objects and create a minimal representation for their data objects, so that they can be tightly coupled with the published research they support. Our goal is to employ automatic techniques as much as possible to generate metadata. Dryad users are also looking for Web 2.0 functionalities, such as, support of annotation and folksonomies, personalization of system features to help with querying or alerting scientists to a new topic (recommender systems), and automatic processing to achieve syntactic interoperability among data sets generated with different software [12].

### 4. Questions about Change

Dryad exemplified change and can help us understand organizing information change and transformation taking place, taking place, drawing from traditional information organization practices, and incorporating newer approaches. DRIADE invites exploration about change, and serves as a probe for considering the applicability of theories of change that might help us understand the transformation of organizing information systems. Questions guiding this initial inquiry are:

- Are aspects of *evolutionary theory* visible that might help explain changes in organizing information systems?
- Are aspect of *cultural diffusion* visible that might explain changes in organizing information systems?

## 5. An Approach to Exploring Questions about Change

Dryad development was initiated in fall 2006, and the repository has been operation for a little over a year. One of the first steps was to develop functional requirements. Formal activities have included an analysis of functionalities found in existing repositories; a stakeholders' workshop/roundtable discussion held in December 2006, which included representatives from the major journals in the field of evolutionary biology, scientists, and representatives of major societies in evolutionary biology; and the recent workshop, "Digital Data Preservation, Sharing, and Discovery: Challenges for Small Science Communities in the Digital Era ([https://www.nescent.org/wg\\_digitaldata/Public:DRIADE\\_Works\\_hop\\_May\\_2007](https://www.nescent.org/wg_digitaldata/Public:DRIADE_Works_hop_May_2007))", which included stakeholders and experts in digital data and organizing information. Information culled from these activities are guiding Dryad's development and show how the theories of evolution and cultural diffusion can explain some of the more novel organizing information practices.

## 6. Reflecting Theories of Change in Dryad

Dryad displays aspects of the theories of evolution and cultural diffusion. Within the construct of evolutionary theory, natural selection, inheritance, and adaptation are applicable, and some time intermingled in results. Examples are presented here:

- **Metadata standards and natural selection, inheritance, and adaptation.** Natural selection intermingled with adaptation may explain, on some level, the use of metadata standards. As indicated above, repository stakeholders have articulated a need and desire for search/retrieval capabilities and recognize that the use of some form of metadata standard and a standardized vocabulary (e.g., a subject thesaurus) will aid help achieve these desired functions. The selection of these features is one of, possibly, *natural selection*—in that they are integral to organizing information. Dryad's functional requirements also include the development of a metadata application profile *inheriting favorable properties* from the Dublin Core, DDI (Data Document Initiative), EML (Ecology Metadata Language), and the PREMIS (PREservation Metadata: Implementation Strategies) data dictionary has been developed to support the repository's first phase. The development of an "application profile" might also be viewed as an *adaptation* in terms of keeping useful feature, and developing a newer and more robust scheme.
- **Annotation and folksonomies as adaptations.** The development of annotation and folksonomy functionalities, where users can provide input both in the form of commentary and tagging, might be explained as *adaptations* of information organization practices that were previously and primarily the responsibility of information professionals.

This activity has become less formalized and less structured, and is increasingly found in information repositories and other types of information environment that support input from users and creators of information, instead of information professionals.

- **The concept of a *work* via natural selection and adaptation.** Although our modern (in the 19/20<sup>th</sup> century sense) information systems have not emphasized the *work* in ways that we might intellectually understand them, the field of library and information science has long recognized the centrality of the idea of a *work* [13]. The concept of a *work* is central to Dryad, and there are two *first class* objects represented in Dryad—the publication and the data objects supporting the publications. The current publication process requires evolutionary biologists to deposit certain data in specialized data repositories (e.g., GenBank and TreeBase), and additional supplementary data in journal repositories (e.g., *American Naturalist* and *Molecular Biology and Evolution*). Tracking the *life-cycle* of digital data objects as they are modified is important for supporting accurate and intelligent reuse of data stored in Dryad. Dryad's support of data object reuse may then be viewed as process of *natural selection*. Dryad objects that are modified as a result of data reuse, and then deposited due to another publication may be considered *adaptations* whose life-cycle might be accurately explained with concepts that form bibliographic relationship taxonomies [14,15,16] to track the life-cycle of data objects, contributing to our understanding of instantiation [17].

Similar to observations related to evolutionary, Dryad exhibits elements of direct and possibly indirect diffusion to explain the theory of cultural diffusion.

- **A team of mixed expertise allowing for direct diffusion.** Dryad's development team includes evolutionary biologists, computer scientists, and information/library scientists. The team of people from different disciplines allows for a direct diffusion of ideas and information about information systems and behaviors or functionalities. We have found that the evolutionary biologists on the Dryad team, and those present at stakeholders meeting, are familiar with the idea of taxonomy in a scientific classificatory sense, for naming species, however, they are neither aware nor comfortable with structured vocabularies underlying information systems, such as lexical semantic thesauri, subject heading lists, and even name-authority files.
- **Specificity and exhaustivity via direct diffusion.** Over the course of the last year, it has become clear that *principled* subject access to Dryad's data objects is critical for resource discovery and for promoting data sharing and reuse. Members of Dryad's development team are working to spreading knowledge of the principles of specificity and exhaustivity via *direct diffusion* through a user tutorial and the development of an "ASK-the-metadata expert" system linking information professionals and scientists.
- **Moving forward via indirect diffusion.** Indirect diffusion implies there is a something that some entity (e.g., a person, country, or an object) may stand between two cultures, although, still, over the course of time, diffusion takes place. *Indirect diffusion* may be taking place as

evolutionary biologists, and scientists in general, are becoming more accustomed to digital repositories and data deposition practices, as they are required to deposit their supplemental data in various journal repositories (e.g., again, *American Naturalist* and *Molecular Biology and Evolution* require the research to submit supplemental data for publications) as part of the publication process. Dryad may be the natural next step for scientists in this domain, given their use of various repositories (e.g., journal repositories, and more specialized repositories, such as Genbank and TreeBase). This idea and the other ideas require more analysis, but it seems likely that indirect diffusion is occurring on multiple levels as scientists, in Dryad's case—evolutionary biologists, work with not only repositories, but general networked technology and information resources daily.

## 5. Conclusion

The paper considered theories of evolution and cultural diffusion for explaining change in organizing information practices. This paper discussed these theories and their value for studying change. A case study focused on the Dryad repository permitted exploration of these theories. This inquiry indicates that the theories of evolution and cultural diffusion are applicable to Dryad, a repository that is reflective of change. Within the construct of evolutionary theory, natural selection, inheritance, and adaptation are applicable, and some time intermingled. Similar to observations related to evolutionary, Dryad exhibits elements of direct and possibly indirect diffusion to explain the theory of cultural diffusion. The application of these theories within Dryad may encourage us to further explore these topics, help us to understand and contextualize change organizing information practices, and identify new significant research questions for advancing our field.

## 2. ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

## 3. REFERENCES

- [1] O'Reilly, T. 2005. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. O'ReillyNet: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [2] Miller, P. 2006. Coming Together around Library 2.0: A Focus for Discussion and a Call to Arms. D-Lib Magazine, 12(4): <http://www.dlib.org/dlib/april06/miller/04miller.html>.
- [3] Darwin, C. 1958. On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life. London: J. Murray.
- [4] Forrest, S. (1993). Genetic Algorithms: Principles of Natural Selection Applied to Computation. Science, 261: 872-878.
- [5] Green S. and Kent, J. P. The Metadata Life Cycle," in MetaNetWork Package 1: Methodology and Tools, ed. Jean-Pierre Kent. 2002, 29-34.



- [6] Schaefer, James M., ed. 1974. Studies in cultural diffusion: Galton's problem: Vol. I, II. Cross-Cultural Research Series.
- [7] Berry, J.W. 1979. Social and Cultural Change. In H.C. Triandis and R. Brislin (Eds.), Handbook of Cross-Cultural Psychology, 5: 211-279. Boston: Allyn and Bacon.
- [8] Bradford, S. C. 1953. Documentation. 2nd ed. London: Crosby Lockwood.
- [9] Rogers, C. E. 1983. Diffusion of Innovations, 5<sup>th</sup>. NY: Simon and Shuster.
- [10] Sonnenwald, Diane H. and Maglaughlin, Kelly L. and Whitton, Mary C. 2001. Using Innovation Diffusion Theory to Guide Collaboration Technology Evaluation: Work in Progress. In Proceedings IEEE International Workshop on Enabling Technologies, pages pp. 114-119.
- [11] Marshall, J. G. 1990. Diffusion of Innovation Theory and End-User Searching. Library and Information Science Research, v12 n1 p55-69 Jan-Mar
- [12] Dube, J., Carrier, S., and Greenberg, J. 2007. DRIADE: a data repository for evolutionary biology. In JCDL2007: Proceedings of the 2007 conference on Digital libraries. p.481, ACM Press, Vancouver, BC, Canada
- [13] Smiraglia, R. P. 2003. The History of "The Work" in the Modern Catalog. CCQ? 553 – 567
- [14] Smiraglia, R. P., and Leazer, G. H. 1999. Derivative Bibliographic Control Relationships: The Word Relationship in a Global Bibliographic Database. Journal of the American Society for Information Science, 50(6): 493–504.
- [15] Tillett, B. 1991. A Taxonomy of Bibliographic Relationships. Library Resources & Technical Services, 35(2), 150-158.
- [16] Tillett, B. B. 1992. Bibliographic Relationships: An Empirical Study of the LC Machine-readable Records. Library Resources & Technical Services, 36(2), 162-88.
- [17] Smiraglia, R. P. 2005. Instantiation: Toward a Theory. In Liwen Vaughan (Ed.) Data, Information, and Knowledge in a Networked World: Proceedings of the Canadian Association for Information Science Annual Conference, June 2-4 2005: <http://www.caais-acsi.ca/search.asp?year=2005>.